Carbon Border Adjustment Mechanisms under Asymmetric Information

Mahaut de Villeneuve Bargemon*

January 2024

Abstract

The paper investigates the optimal design of a Carbon Border Adjustment Mechanism to mitigate carbon leakage created by the relocation of domestic firms to unregulated countries. Such regulation was proposed by the European Commission in July 2021. However, this system relies on firms self-reporting their emissions to determine the carbon tax to be paid on traded goods. Authorities are unlikely to be able to verify the veracity of such information, inducing firms to falsely report low emissions. With a theoretical model in which the regulator has incomplete information about firms' levels of pollution, we show that he must design a non-linear tax structure to distinguish between dirty and clean firms. In this setup, the "good type" (*i.e.* clean firm) is also the less cost-efficient, which constrains the regulator to reward dirty firms and impose a decrease in clean production compared to the optimal solution. This feature also uncovers an important friction between incentives: when firms can choose their technology of production, it is impossible for the regulator to reconcile both the incentive to become clean and a truth-revealing tax structure.

*PhD Candidate in Economics - ESSEC Business School, CY Cergy Paris University, ThEMA
Contact: mahaut.devilleneuvebargemon@essec.edu // mahaut.de-villeneuve1@cyu.fr
Acknowledgments: I wish to sincerely thank my supervisor, Professor Wilfried Sand-Zantman, for his guidance, advice, patience and support. I am also grateful to Professor Anastasios Dosis from ESSEC for his comments, questions and suggestions during seminars. Many thanks to Professors Aude Pommeret, Katheline Schubert, and Carolyn Fischer for their valuable comments and for the crucial questions they raised during the 2023 EAERE-ETH European Winter School.

1 Introduction

Research Question. Global warming is induced by the excess of greenhouse gases emissions into the atmosphere from human activities. The environment can be considered as a global public good and therefore suffers from the *"greatest and widest-ranging market failure ever seen"* (Stern, 2007). The optimal solution to regulate carbon emissions would be the implementation of a global carbon price, but efforts towards such regulation have fallen short. The emergence of regional regulations to pursue climate ambitions means that, while regulated domestic firms pay a carbon price, foreign firms located in more lenient countries do not. A widely-discussed solution for regulating these outside firms is the implementation of a Carbon Border Adjustment Mechanism (CBAM), which takes the form of a border tax imposed on the carbon content of entering goods. In this paper, we are interested in the design of a CBAM when the emissions of firms located abroad are unknown to the regulator. In other words, what is the optimal CBAM regulation under incomplete information about firms' technologies?

Context. Although global negotiations on climate mitigation have led to the adoption of important international environmental agreements such as the Kyoto Protocol (adopted in 1997) and the Paris agreement (2015), these agreements remain non-binding and never led to the creation of a unique carbon price. Therefore, countries face incentives to behave as free-riders (Nordhaus, 2015). As a response to the inability of the global community to decide on a unique regulation, subglobal regulations have emerged. For instance, the European Union created the EU Emissions Trading System (EU ETS) in 2005, the largest market for emission allowances. These types of regulations enabled the imposition of a domestic carbon price, but created a new issue: the issue of carbon leakage. Carbon leakage is defined as an increase in foreign emissions following a reduction in domestic emissions (European Commission, 2015). Therefore, regulating foreign emissions is a crucial challenge for regional regulations as the EU ETS, because not doing so can compromise efforts and even result in increased emissions globally (Monjon & Quirion, 2011). Empirical studies estimate the carbon leakage rates between 10 percent and 30 percent (Bohringer et al., 2018). The creation of a CBAM is intended to tackle this issue. The European Commission voted the creation of such a mechanism in 2021 as part of its *Fit for 55* package. However, the design of a CBAM presents many challenges. Among them, the inability of governments to access information about the carbon

content of imported goods. When the regulation relies on self-reporting to retrieve this information, foreign firms with carbon-intensive technologies face incentives to under-report their emissions in order to pay a lower tax.

Model Overview and Main Results. This paper proposes a simple theoretical model of regulation in which a social planner wants to design a CBAM without having information about foreign firms' emissions. We consider firms with two different technologies: dirty firms, which have a lower cost of production but pollute more, and clean firms, which do not pollute but are less cost-efficient. The regulator has complete information about emissions when firms are located inside the country. However, he does not access this information anymore when firms relocate their production abroad, which is the issue of interest.
The benchmark case under complete information gives the welfare-maximizing optimum. At this optimum, the regulation is designed such that the environmental negative externality created by dirty firms is internalized, and tax revenue is maximized. However, we show that the welfare-maximizing regulation is not applicable to firms located abroad, because of incomplete information about their technologies. The design of a second-best regulation therefore needs to ensure that each firm is willing to declare its true technology to the regulator by picking the tax designed for its type. This results in distortions of quantities and taxes charged on firms compared to the complete information benchmark. Moreover, adding the WTO concerns in this analysis entails further distortions.

Related Literature and Contribution. This work relates to two main strands of literature.
First, it aims to contribute to the literature on Carbon Border Adjustment Mechanisms, and more generally on the use of border tax structures to address carbon leakage. One of the first papers to consider tax structures (production, consumption and trade taxation) as a way to correct international externalities (such as pollution) and maximize social welfare is Markusen (1975). Hoel (1996) extends this framework to allow for more than two countries and goods traded, and to consider differentiation across sectors. Both of these papers consider a model with complete information about firms' costs and damages. The use of such tax structures for the purpose of a CBAM has gained prominence in the literature in recent years. Some studies such as Cosbey et al. (2019) provide an overview of the legal and economic issues at stake in order to offer guidelines on its design and implementation. They

acknowledge and discuss the difficulties raised by the presence of incomplete information, but do not derive a specific model to deal with this issue. This strand of literature has undertaken a quantification of the amount of carbon leakage in order to demonstrate the importance of this issue. It is estimated to lie between 5 and 30 percent for industrialized countries (Bohringer, Balestreri and Rutherford 2012). It also justifies the adoption of a CBAM to tackle this problem (Monjon & Quirion, 2011). To summarize, there are concerns regarding the assessment of the carbon content of imported products, but most papers rely on the use of default values as a way out. The present paper intends to address this issue by incorporating incomplete information in the design of a CBAM.

This paper also aims to contribute to the theoretical literature on pollution control under incomplete information. In Kwerel (1977), a regulator wants to regulate polluting firms without knowing their characteristics. In their model, firms are heterogeneous in cost of production but they all impose the same environmental damage, which results in an optimal regulation with a unique price for all firms. In the present paper, we allow for firms with different costs of production to be more or less polluting. Spulber (1988) also considers a model in which a social planner wants to regulate heterogeneous polluting firms which have private information about their costs. While this paper assumes that the regulator assigns firms emission levels and payments for damage after firms send a message (signal), the present paper intends to design a model with one-round communication only (no signaling, the regulator does not have any information about firms when designing the optimal policy). The paper by Dasgupta, Hammond and Maskin (1980) improves Kwerel (1977) by designing a model in which truth-telling is dominant strategy for all firms, and not only Nash equilibrium. They consider both one-round communication (the social planner chooses the regulation without any information) and two-round communication (signaling). This paper derives a modified VCG mechanism, which applicability to policy-making remains vague. We enrich this approach by designing a different kind of model (Principal-Agent style) by adding the concerns with WTO rules compatibility, and by considering endogenous investment in the clean technology by firms. Finally, we assume that clean firms, which are the "good type" from an environmental damage perspective, are the less cost-efficient type. This results in the necessity of rewarding the polluting firms to acquire information. Such assumption differs from traditional models in which the less costly type is always the good type.

Structure of the Paper. The structure of this paper is as follows. First, we build a benchmark case in which we consider the choice of a tax structure to regulate domestic firms, whose technologies are known by the regulator (Section 2). In Section 3, we turn to the problem of interest and derive the CBAM chosen by the regulator for firms located abroad, assuming that he does not have any information about their emissions, and that he needs to induce them to voluntary declare their true technology to him. Then, Section 3 explores the compatibility of this regulation with WTO rules, and derives a policy that would be more likely to be in line with these requirements. In Section 5, we allow for firms to choose whether to invest in the clean technology, and we discuss the tension between inducing firms to adopt such "good" but costly behavior and the truthful report of technology. We demonstrate that the two dimensions cannot be reconciled. Finally, Section 6 concludes the paper.

2 Model and Benchmark

Consider a market with consumers, firms, and a regulator.

Consumers. For simplification, assume that demand is perfectly elastic. As a consequence, at a given price, consumer surplus is zero.

Firms. The market is composed of a mass one of perfectly competitive and profit-maximizing firms. There are two technologies of production: firms can either be clean (C) with probability $\lambda \in [0, 1]$, or dirty (D) with probability $(1 - \lambda)$. Each firm $i \in \{C, D\}$ produces a quantity $q_i \geq 0$ of the homogeneous good, and sells this good at a fixed price $\overline{p} = 1$. Profits from production depend on: (*i*) the revenue from selling the good, (*ii*) the cost of production, assumed to be increasing and convex in the quantity produced, and (*iii*) the tax payment to the social planner. Thus, assume that profits write $\pi_i(q_i, t_i) = q_i - \frac{1}{2\theta_i} q_i^2 - t_i q_i$ where $\theta_i \geq 1$ is the firm's type, and $T_i = t_i q_i \geq 0$ is the total tax on production paid to the regulator. Assume that clean firms have a higher cost of production than dirty firms: $\theta_C < \theta_D$. Indeed, at least in the short term, clean firms might have higher costs of production arising from acquiring renewable energy and greener raw materials, especially in a framework where they are not able to differentiate their product from "brown" products.[1] Clean

[1]For insights on the possibility of reducing costs by going green that is ruled out in this paper, refer to the works of Porter and van der Linde, 1995.

firms do not emit carbon dioxide, therefore they do not impose any environmental damage on society. However, dirty firms create a damage $\gamma > 0$ per unit produced. The total environmental damage created by a dirty firm is thus equal to γq_D.

Regulator. The regulator is a benevolent social planner who maximizes society's welfare. He can choose to tax firms. He values firms' profits, and the extraction of money from firms through the carbon tax (tax revenue), with weight $\beta \in [0,1]$ (Laffont and Tirole, 1996).[2] He dislikes the environmental damage imposed by dirty firms. Formally, we can therefore write the welfare function as

$$W = \lambda\Big(\pi_C(q_C, t_C) + (1+\beta)t_C q_C\Big) + (1-\lambda)\Big(\pi_D(q_D, t_D) - \gamma q_D + (1+\beta)t_D q_D\Big) \tag{1}$$

2.1 Laissez-Faire

First consider the situation in which there is no regulation chosen by the social planner: $t_C = t_D = 0$. In this case, each firm chooses to produce the quantity that maximizes its individual profits $\pi_i(q_i) = q_i - \frac{1}{2\theta_i}q_i^2$. Thus, the *laissez-faire* quantity produced by each type of firm is $q_i^* = \theta_i \; \forall i \in \{C, D\}$, and the resulting profits are equal to $\pi_i^*(q_i^*) = \frac{1}{2}\theta_i$. Dirty firms produce more than clean firms because they are more cost-efficient, and therefore they make higher profits. They do not account for the negative externality created by their emissions when maximizing their own profit function, imposing a total damage on society equal to $\gamma\theta_D$. In this case, total welfare is equal to $W^* = \frac{1}{2}\lambda q_C^* + (1-\lambda)q_D^*\Big(\frac{1}{2} - \gamma\Big)$.

2.2 Regulation under Complete Information

Now, how can the *laissez-faire* situation be improved with the intervention of a regulator that maximizes society's welfare? In this subsection, we derive the welfare-maximizing regulation[3] consisting of quantities to be produced by each type of firm and taxes to be paid. Assume that all firms are domestic (the regulator cares about their profits) and located inside the country. In this case, the regulator is able to acquire perfect information about firms' emissions. In other words, the regulator

[2] This is motivated by the double dividend hypothesis: environmental taxes can both reduce environmental damages and finance reductions in other types of taxes which are sources of distortions.

[3] We denote the elements of the welfare-maximizing regulation with the superscript *FB*, which stands for "first-best", to follow the usual notations in contract theory.

can perfectly verify the technology of production of firms: he has complete information about firms' characteristics. He is therefore able to design an environmental regulation based on this observation. To illustrate this situation, consider how EU authorities can verify the European producers' emissions, covered by the EU ETS. The regulator chooses the quantities produced by each type of firm, and the per-unit taxes charged, in order to maximize society's welfare (1).[4] Replacing profits and per-unit taxes by their expressions, we can rewrite:

$$\begin{aligned} W &= \lambda\Big(\pi_C(q_C, t_C) + (1+\beta)t_C q_C\Big) + (1-\lambda)\Big(\pi_D(q_D, t_D) - \gamma q_D + (1+\beta)t_D q_D\Big) \\ &= \lambda\Big((1+\beta)\Big(q_C - \frac{1}{2\theta_C}q_C^2\Big) - \beta\pi_C\Big) + (1-\lambda)\Big((1+\beta)\Big(q_D - \frac{1}{2\theta_D}q_D^2\Big) - \gamma q_D - \beta\pi_D\Big) \quad (1.2) \end{aligned}$$

Writing the social welfare in terms of quantities and profits as in (1.2) gives some intuition about the way society values taxation comes into play: gross profits (before taxation) made by firms are valued with a weight superior to one because they represent the amount that may be retrieved by the regulator with a tax. On the other hand, net profits enter negatively in the social welfare, showing that society is willing to tax firms as high as possible.
Formally, the regulator solves:

$$\begin{aligned} &\max_{q_D, q_C, \pi_C, \pi_D} (1.2) \\ \text{s.t.} \quad &\pi_C \geq 0 \quad \text{and} \quad \pi_D \geq 0 \end{aligned}$$

Therefore, the regulator chooses the highest possible tax rate such that firms make non-negative profits. As a consequence, he maintains profits of both types of firms at zero ($\pi_C^{FB} = \pi_D^{FB} = 0$). He extracts the entire gross profits using taxation, with $t_i^{FB} = 1 - \frac{1}{2\theta_i}q_i$. He lets clean firms produce their *laissez-faire* quantity $q_C^{FB} = q_C^* = \theta_C$. However, he requires dirty firms to produce a lower quantity than their profit-maximizing production: $q_D^{FB} = \theta_D\Big(1 - \frac{\gamma}{1+\beta}\Big)$. To ensure non-negative production, we impose the following assumption: $\gamma < 1 + \beta$.
In this situation, social welfare will be $W^{FB} = \frac{1}{2}\lambda(1+\beta)q_C^{FB} + \frac{1}{2}(1-\lambda)(1+\beta-\gamma)q_D^{FB}$.
The following proposition summarizes the welfare-maximizing regulation under complete information.

Proposition 1 (Benchmark under complete information)**.** *Suppose all firms are located within the country, implying complete information about their technologies. The regulator*

[4] By construction, the maximization program of the regulator is always concave in both q_C and q_D for all acceptable values of the model parameters.

chooses a regulation in the form of a tax structure $\{(t_C, q_C); (t_D, q_D)\}$ *specifying, depending on technology, a per-unit tax on production and a quantity cap to maximize social welfare, inducing:*

(i) Dirty firms to produce less than in laissez-faire; and clean firms to produce their laissez-faire quantity.

(ii) Taxes set such that all firms' profits are equal to zero, to maximize tax revenue.

Proof. See the Appendix. □

The regulator lets clean firms produce as in *laissez-faire* because they do not create any damage. However, dirty firms are required to produce a quantity lower than their *laissez-faire* production in order to internalize the negative externality from emissions.

The quantity produced by the dirty type is increasing in its type $\left(\frac{\partial q_D^{FB}}{\partial \theta_D} > 0\right)$, meaning that a more cost-efficient firm is allowed to produce more, as the regulator values gross profits. The results show the existence a trade-off between environmental protection and the extraction of profits through taxation. Indeed, the quantity produced by dirty firms is decreasing in the marginal damage of emissions $\left(\frac{\partial q_D^{FB}}{\partial \gamma} < 0\right)$, because it is optimal to pollute less when the environmental damage is higher. However, it is increasing in the value attributed by society to the extraction of profits $\left(\frac{\partial q_D^{FB}}{\partial \beta} > 0\right)$. Indeed, when the social planner highly values tax revenue (high β), he is willing to let dirty firms produce more in order to extract more. Thus, parameter β illustrates the willingness of the regulator to increase dirty firms' production above the level that would be required if we were to mitigate the environmental damage only (in other words, above the "Pigouvian" level of production).

Quite directly, the welfare attained under this regulation is higher than in *laissez-faire*. Under the zero-profit condition maintained by the social planner, note that per unit taxes and total taxes move in opposite directions when quantity varies. This happens because profits are increasing and concave, as costs of production are increasing convex. Indeed, per unit taxes ($t_i = 1 - \frac{1}{2\theta_i} q_i$) are decreasing in the quantity produced, q_i. On the other hand, total taxes ($T_i = q_i - \frac{1}{2\theta_i} q_i^2$) are increasing in q_i.[5] This describes how taxes imposed on clean and dirty firms will vary when quantities vary following a variation in exogenous parameters (β, γ, θ_i).

The per unit tax imposed on dirty firms is equal to $t_D^{FB} = \frac{1}{2}(1 + \frac{\gamma}{1+\beta})$, which is greater

[5] This is true for $q_i < \theta_i$, which is always the case here.

than the per unit tax imposed on clean firms: $t_C^{FB} = \frac{1}{2}$. If production did not harm society with the creation of an environmental damage ($\gamma = 0$), per unit tax on clean firms and dirty firms would be identical. However, as soon as there exists a positive environmental damage $\gamma > 0$, the regulator requires dirty firms to produce less than their profit-maximizing quantity, to internalize the negative effect. The decrease in quantity translates to an increase in per unit tax. A positive value of β attenuates this effect.

There is no issue of implementation when considering domestic firms, as the regulator is able to perfectly verify firms' technologies. Thus, firms are directly required, depending on their type θ_i, to produce the welfare-maximizing quantity and pay the corresponding tax to the regulator.

The following graph illustrates the regulation chosen by the social planner for domestic firms, in complete information. For interpretation purposes, the imposition of a quantity to be produced can be regarded as a cap that firms are not allowed to exceed on this market. Firms being profit-maximizing, they will choose to produce this maximum quantity exactly (all firms will bunch on the blue points). In other words, from zero production up to $q_i = q_C^{FB}$, firms are required to pay a total tax equal to $\frac{1}{2}q_i$ (green area on the graph). From $q_i = q_C^{FB}$ up to $q_i = q_D^{FB}$, firms are required to pay $\frac{1}{2}q_i(1 + \frac{\gamma}{1+\beta})$ (yellow area). This is represented by the non-linear red function.

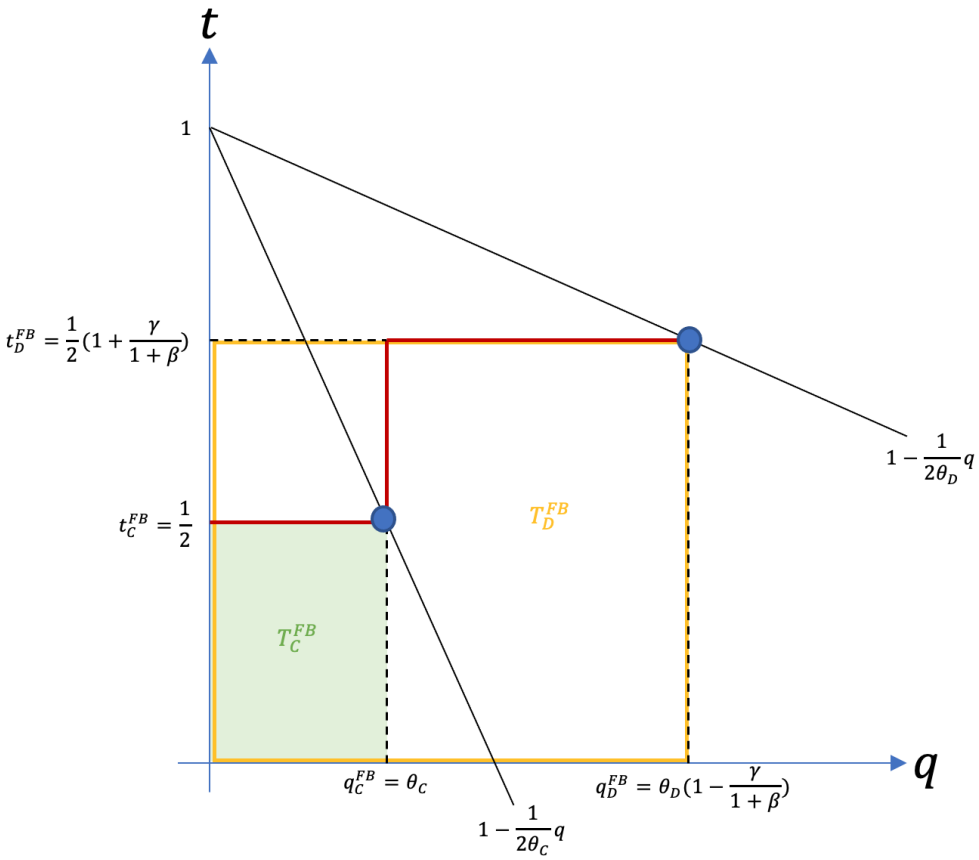


Figure 1: Optimal Regulation of Domestic Firms under Complete Information

3 Regulation of Firms Relocated in Foreign Countries

3.1 Carbon Leakage, Rationale for a CBAM & Asymmetric Information

In the previous section, we derived the optimal regulation chosen by the regulator for domestic firms when they are located inside the country. Now, the aim of this paper is to investigate the design of a regulation when firms try to escape the tax by relocating their production plants outside the country. This is typically the issue of carbon leakage. A Carbon Border Adjustment Mechanism is a type of regulation seeking to put a price on the carbon emitted during the production of goods that are entering the country (or the region, in the case of the EU). It aims to ensure a level playing field between firms located inside and outside the area. However, we assume that the regulator cannot observe the characteristics of the firms located

abroad: he does not access information about firms' emissions (type θ_i is private information). This is a crucial concern for the design of a CBAM, because it is very difficult to get information about the carbon emissions of firms located in foreign countries. The European Commission's proposal relies on a declarative system, in which firms self-report their emissions. This implies that firms may "lie" about their technologies and under-report their emissions when it is profitable. The use of audits to deter such behavior was mentioned in the proposal, but remains vague and would be very costly, if not unfeasible. Also note that in the case of the absence of declaration, firms will be assigned "default values" based on the average emission intensity of the worst 10 percent performing EU producers. On top of the fact that this goes against the logic of carbon pricing, it does not guarantee an accurate reflection of the carbon emissions associated with the imported goods.[6]

3.2 Solving for the Optimal CBAM under Asymmetric Information

The problem of incentives. For simplification, assume that all firms (clean and dirty, total mass of one) have relocated their production plants outside of the country, produce and export the entire quantity to the domestic country. The regulator does not know θ_i. He designs a tax structure and lets firms pick the tax (and associated quantity) themselves in $\{(q_C, t_C); (q_D, t_D)\}$. In this case, offering the regulation derived in complete information from the previous section (Proposition 1) does not allow to reach maximum welfare. Indeed, when faced with the first-best regulation, dirty firms prefer not to choose according to their actual technology, but mimic the clean types (by choosing (q_C, t_C)) to earn a positive profits: $\pi_D(q_C^{FB}, t_C^{FB}) > 0$. In this case, the regulator is unable to distinguish between the two technologies. Clean firms however choose accordingly (because $\pi_C(q_D^{FB}, t_D^{FB}) < 0$).

Solving the problem. Because it is more profitable for dirty firms to "lie" about their technology, the regulator cannot tax the firms as in complete information if he wants to distinguish between the different firms, so he needs to design the regulation differently. In particular, to ensure that dirty firms choose the correct tax, he will need to let them earn some profits.[7]

[6] Moreover, studies call for a differentiation of benchmark values among exporters but this raises two issues: (i) data for country-specific default values may not be available, (ii) exporter-specific benchmarks might violate GATT's Most Favoured Nation principle (Cosbey et al., 2019).

[7] In contract theory, this is usually called information rent.

Formally, the regulator chooses a tax structure $\{(q_C, t_C); (q_D, t_D)\}$ to maximize the social welfare function (1.2) subject to

$$q_C - \frac{1}{2\theta_C} q_C^2 - t_C q_C \geq q_D - \frac{1}{2\theta_C} q_D^2 - t_D q_D \tag{2}$$

$$q_D - \frac{1}{2\theta_D} q_D^2 - t_D q_D \geq q_C - \frac{1}{2\theta_D} q_C^2 - t_C q_C \tag{3}$$

$$\pi_C(q_C, t_C) \geq 0 \tag{4}$$

$$\pi_D(q_D, t_D) \geq 0 \tag{5}$$

The first two inequalities (2) and (3) ensure that each type of firm (clean and dirty respectively) makes higher profits when choosing the correct tax (intended for its technology), compared to the profits it would earn when lying.
The latter two constraints (4) and (5) ensure that none of the firms make negative profits, and therefore both types are willing to participate in the market.[8]
The problem can be simplified to maximizing the social welfare function subject to $\pi_D^{SB} = q_C^{SB^2}\left(\frac{1}{2\theta_C} - \frac{1}{2\theta_D}\right)$ and $\pi_C^{SB} = 0$.[9] Solving the problem gives the following solutions, denoted with the superscript SB standing for "second-best", to remain consistent with the previous section's notation. The total tax imposed on clean firms is set such that their profits are equal to zero: $T_C^{SB} = q_C^{SB} - \frac{1}{2\theta_C} q_C^{SB^2}$. Thus, the per-unit tax on clean production is equal to $t_C^{SB} = 1 - \frac{1}{2\theta_C} q_C^{SB}$. On the other hand, the total tax imposed on dirty firms is equal to $T_D^{SB} = T_D^{FB} - \pi_D^{SB}$. Therefore, the per-unit tax writes $t_D^{SB} = t_D^{FB} - \frac{\pi_D^{SB}}{q_D^{SB}}$, and not only depends on the quantity produced by dirty firms, but also on the quantity produced by clean firms, which affects the profits that dirty firms earn. While the quantity produced by dirty firms is identical to the first-best regulation quantity ($q_D^{SB} = q_D^{FB}$), the quantity produced by clean firms is equal to $q_C^{SB} = \varphi q_C^{FB}$ where $\varphi = \frac{\theta_D \lambda(1+\beta)}{\theta_D(\lambda+\beta) - \theta_C \beta(1-\lambda)} < 1$. As a result, dirty firms earn positive profits $\pi_D^{SB} > 0$.
The following proposition describes the second-best regulation chosen by the regulator under incomplete information.

Proposition 2 (Second-best regulation under incomplete information about firms located abroad)**.** *Suppose all firms are located abroad, implying incomplete information about their technologies. The regulator chooses a regulation (CBAM) in the form of a*

[8] In other words, constraints (2) and (3) are incentive compatibility constraints; while (4) and (5) are participation constraints.

[9] Concavity of the program in q_D is always satisfied. Concavity of the program in q_C is satisfied for $(1-\lambda)\beta\theta_C \leq (\lambda+\beta)\theta_D$. Details on the simplification of the problem can be found in the Appendix.

tax structure $\{(t_C, q_C); (t_D, q_D)\}$ *specifying per-unit tax on production and quantity cap to maximize social welfare such that firms reveal their true technology, inducing:*

(i) No distortion of the quantity produced by dirty firms compared to the first-best regulation, but a lower per-unit tax paid by these firms, allowing them to earn positive profits ("information rent").

(ii) A downward distortion of the quantity produced by clean firms compared to the first-best. Taxes on production are still set such that clean firm profits are equal to zero.

(iii) In general, lower total taxes for both types of firms.

Proof. See the Appendix. □

The regulator decreases the quantity that clean firms are allowed to produce in order to ensure that dirty firms behave correctly. Indeed, by distorting production in that way, it makes it less attractive for dirty firms to lie. Moreover, the quantity produced by clean firms is decreasing in θ_D. A more cost-efficient dirty firm has more incentive to lie, so the quantity required for the clean type must be decreased for dirty firms to act according to their true technology.

Total taxes are non-linear and increasing in quantities.[10] The total tax imposed on clean firms is lower than in the first-best regulation because clean firms' profits are maintained at zero but the quantity produced is decreased. The total tax imposed on dirty firms is also lower in second-best compared to the first-best regulation because they make positive profits, while the quantity they produce is identical. Because $q_D > q_C$ (implied by (2) and (3)), the total tax imposed on the dirty type is always greater than the total tax imposed on the clean type: $T_D^{SB} > T_C^{SB}$. Indeed, the dirty type is the most cost-efficient at producing the good (lower cost of production) and pays a higher total tax but is willing to do so in order to produce more.

As mentioned in the previous section, total tax and per-unit tax respond to changes in quantities in opposite directions. Thus, the per-unit tax paid by clean firms increases compared to the first-best solution. On the side of dirty firms, the quantity produced did not vary but the regulator chooses a lower per-unit tax than what is optimal in order to provide the incentive to disclose the truth.

In this model, clean firms do not impose a damage on society (they are the "good" type with respect to the environment), but they are less cost-efficient than dirty

[10] T_D^{SB} is increasing in q_D as long as $q_D < 2\theta_D$, which holds. T_C^{SB} is increasing in q_C as long as $q_C < 2\theta_C$, which is satisfied as well.

firms. As a consequence, under incomplete information, the regulator is forced to reward dirty firms in order to acquire information (by letting them earn some profit) and hurt clean firms (by decreasing their production), although he would ideally like to let them produce their profit-maximizing quantity. These results show how information asymmetries may undermine environmental efforts when being green is more costly.

It is intuitive and straightforward to demonstrate that the resulting social welfare is lower than under complete information: $W^{SB} < W^{FB}$. On the one hand, the part of welfare emerging from the activity of clean firms decreases via the decrease in quantity produced. On the other hand, the part of welfare emerging from the activity of dirty firms also decreases because the quantity produced remains constant with respect to the first-best, while the total tax imposed on dirty firms decreases. This decrease in total social welfare does not come from greater environmental damage compared to the benchmark: as mentioned previously, the quantity produced by dirty firms being identical as in the first-best solution, therefore the total environmental damage as well.

On the contrary, it is worth noting that the results leave room for the possibility of achieving a lower environmental damage. If we assume that clean firms emit even a slightly positive amount of carbon dioxide when they produce, inducing them to impose an environmental damage on society (the marginal damage imposed by clean firms should be lower than γ), decreasing their production for incentive purposes will decrease the overall total environmental damage, although not targeting the most polluting producers on the market.

The following graph gives an illustration of the regulation chosen by the regulator for firms located abroad, under incomplete information. The imposition of a specific quantity to produce can be viewed as in the last section, *i.e.* as a cap on the imported quantity of products. Being profit-maximizing, clean firms will choose exactly q_C^{SB} and dirty firms will produce q_D^{SB}, bunching on the two blue points in the figure. The per-unit tax schedule (red line) is still non-linear.

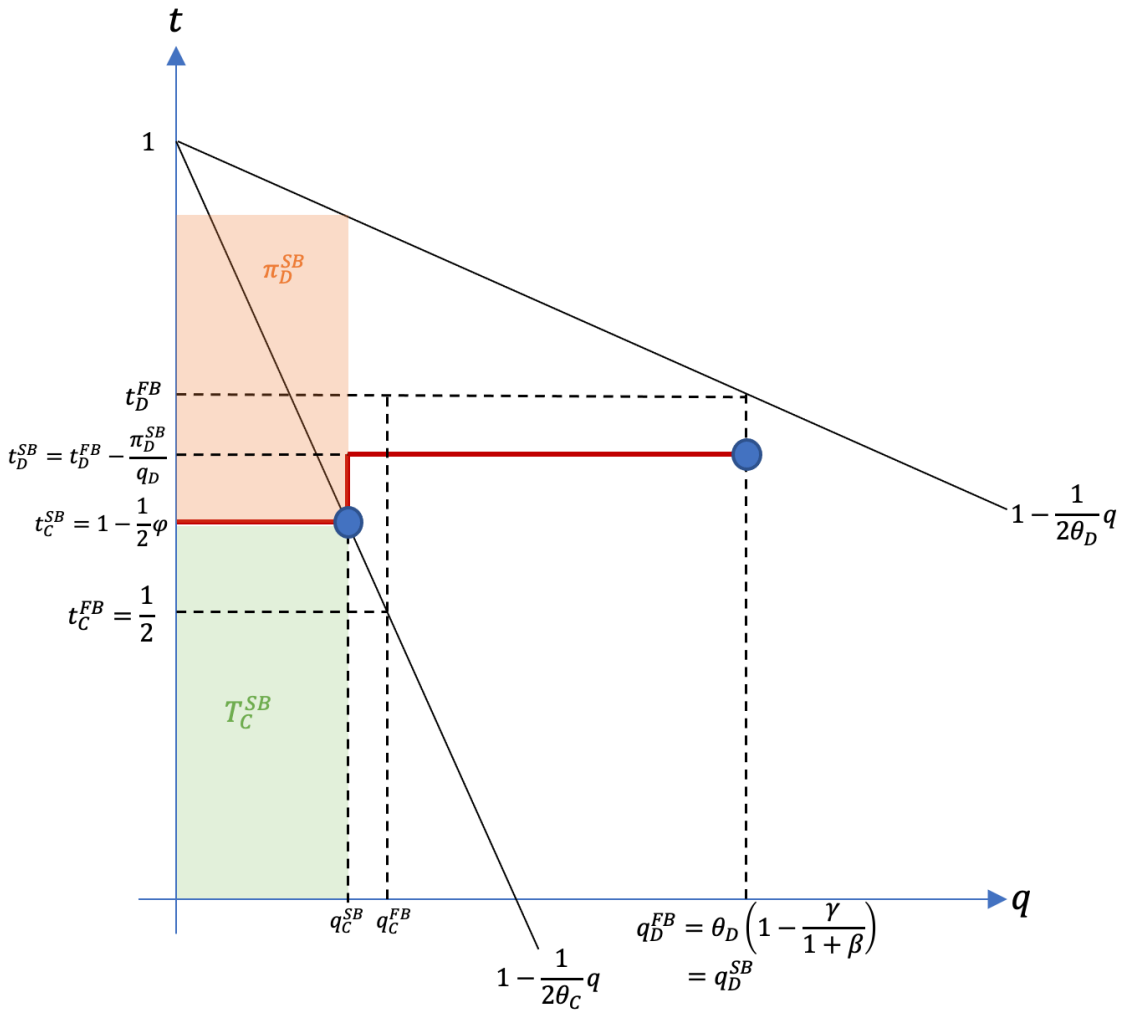


Figure 2: Regulation of Firms Located Abroad under Incomplete Information

4 WTO Compatibility

The purpose of designing an environmental policy with taxation is to make the polluting firms pay for their emissions. However, in this model, the regulator also tries to extract profits from firms. Applied to imported goods, this framework is not likely to be compatible with the World Trade Organization rules. A mechanism that would be more likely to satisfy the WTO requirements would target only pollution (Cosbey et al., 2019 ; Pauwelyn, 2013). For instance, consider the design of a tax on imported goods that must not exceed the Pigouvian tax level.[11] In this case, the

[11] We define the Pigouvian per-unit tax rate as being equal to the marginal environmental damage.

regulator would not be able to charge clean firms (as they create no environmental damage), *i.e.* we must have $t_C^W = 0$; and should choose the per-unit tax charged on dirty firms to satisfy $t_D^W \leq \gamma$. At least the first one is violated in the second-best regulation, so it is relevant and must be imposed with equality. It is possible to show that the second constraint is not necessarily violated for t_D^{SB}: for all values of $\beta \in [0, 1]$, there exists a range of acceptable (under the model's assumptions) values of γ that satisfy the WTO requirement. Thus, this constraint will be included as an inequality in the social planner's problem, and not imposed directly with equality.[12] Note that, clean firms not being subject to a tax anymore, they will make positive profits as long as $q_C^W \in [0; 2\theta_C]$.

Solving the problem. To design a regulation targeted to foreign firms with no information about their technology and that is compatible with the WTO rules, the regulator must maximize the welfare function (1.2) subject to

$$\pi_C \geq q_D - \frac{1}{2\theta_C} q_D^2 - t_D q_D \tag{6}$$

$$\pi_D \geq q_C - \frac{1}{2\theta_D} q_C^2 \tag{7}$$

$$\pi_C \geq 0 \tag{8}$$

$$\pi_D \geq 0 \tag{9}$$

$$t_C = 0 \Rightarrow \pi_C = q_C - \frac{1}{2\theta_C} q_C^2 \tag{10}$$

$$t_D \leq \gamma \tag{11}$$

Similarly to Section 3, the first two constraints ensure that both types of firms choose correctly when facing the regulation, and do not lie about their technology. Clean firms are not likely to lie, so (6) can be ignored and checked ex-post. Inequalities (8) and (9) ensure the participation of both types. Obviously, constraints (10) and (11) ensure compatibility with the WTO rules, as discussed above. The problem can be simplified to maximizing the welfare function (1.2) subject to $\pi_D^W = q_C^W\left(1 - \frac{1}{2\theta_D} q_C^W\right)$ and $\pi_C^W = q_C^W\left(1 - \frac{1}{2\theta_C} q_C^W\right)$.[13] The regulator chooses the same quantity to be produced by dirty firms as in first-best and second-best regulations: $q_D = \theta_D\left(1 - \frac{\gamma}{1+\beta}\right)$. He sets the quantity to be produced by clean firms lower than in the second-best regulation:

[12] Refer to the Appendix for an intuition on this matter.

[13] Concavity of the program in q_D is always satisfied. Concavity of the program in q_C is assured for $\lambda\theta_D > (1 - \lambda)\beta\theta_C$. Details on the simplification of the problem can be found in the Appendix.

$q_C^W = \eta q_C^{FB}$ where $\eta = \frac{\theta_D(\lambda-(1-\lambda)\beta)}{\lambda\theta_D-(1-\lambda)\beta\theta_D} < \varphi < 1$. He chooses to impose the following tax on dirty firms: $T_D^W = T_D^{FB} - \pi_D^W$, which is equivalent to imposing a per-unit tax on production equal to $t_D^W = t_D^{FB} - \frac{\pi_D^W}{q_D}$.
The following proposition describes the regulation offered to the firms located in foreign unregulated countries under incomplete information about their technologies and to comply with the Pigouvian taxation constraints.

Proposition 3 (WTO-compatible regulation under incomplete information)**.** *Suppose all firms are located abroad, implying incomplete information about their technologies. Also suppose that taxing firms more than the Pigouvian level violates WTO rules. The regulator chooses a regulation (CBAM) in the form of a tax structure* $\{(t_C, q_C); (t_D, q_D)\}$ *to maximize social welfare such that firms reveal their true technology and to ensure compliance with the WTO, inducing:*

(i) No distortion of the quantity produced by dirty firms compared to the first-best and second-best regulations.

(ii) An additional downward distortion of the quantity produced by clean firms compared to the second-best regulation.

(iii) An increase or a decrease of dirty firms' profits depending on the share of clean firms λ *and the value attributed to tax revenue* β*. Positive profits earned by clean firms.*

(iv) Zero taxation imposed on clean firms. A higher or a lower total tax imposed on dirty firms depending on the change in profits.

Proof. See the Appendix. □

The necessity to further distort the production of clean firms arises from the inability to tax these firms. Indeed, the absence of taxation when declaring θ_C makes it even more attractive for a dirty firm of type θ_D to lie about its technology. To ensure that dirty firms do not mimic the clean technology, the quantity that clean firms are allowed to produce must be reduced.
Profits made by dirty firms increase with the inclusion of the Pigouvian taxation constraint compared to the second-best regulation when β is small enough (tends towards 0) and/or λ is high enough (tends towards one). On the contrary, dirty profits decrease with the WTO-compatible regulation compared to the second-best when β is high enough and/or λ is small enough. As a consequence, the total tax charged on dirty firms decreases in the first case ($T_D^W < T_D^{SB}$), and increases in the

second ($T_D^W > T_D^{SB}$). Because the quantity produced by dirty firms stays constant, an increase in their profits under the WTO constraint implies a lower per-unit tax with respect to the second-best, while an decrease implies a higher per-unit tax.

Note that, in order to satisfy concavity, it is more likely that λ is high and/or β small. The intuition is the following: when λ is high (there are a lot of clean firms) and/or β is small (the regulator is not willing to highly distort quantities), the regulator only slightly decreases the quantity produced by clean firms (small downward distortion). But then, to make sure that dirty firms still choose according to their technology, the regulator must allow for them to make higher profits. Take the extreme case where $\lambda = 1$ and $\beta = 0$. In this case, the quantity produced by clean firms under the second-best regulation and under the WTO-compatible regulation is identical ($q_C^{SB} = q_C^W = \theta_C$) ; and profits made by dirty firms from the informational advantage increase in the WTO-compatible regulation with respect to the second-best ($\pi_D^W > \pi_D^{SB}$).

Similar to the second-best tax discussion, it is possible to show that, for all values of β, there exists a range of acceptable values of the marginal damage γ such that the WTO constraint $t_D^W \leq \gamma$ is satisfied.[14]

The following graph illustrates the regulation chosen by the social planner for foreign firms and to comply with the Pigouvian taxation rule, in the case where $\pi_D^W > \pi_D^{SB}$. The red line represents the non-linear per-unit tax schedule, and firms still bunch on the two blue points because of profit-maximization. Both types of firms make positive profits (green and orange areas), with clean profits unambiguously lower than dirty profits.

[14] Refer to the Appendix for intuition on that matter.

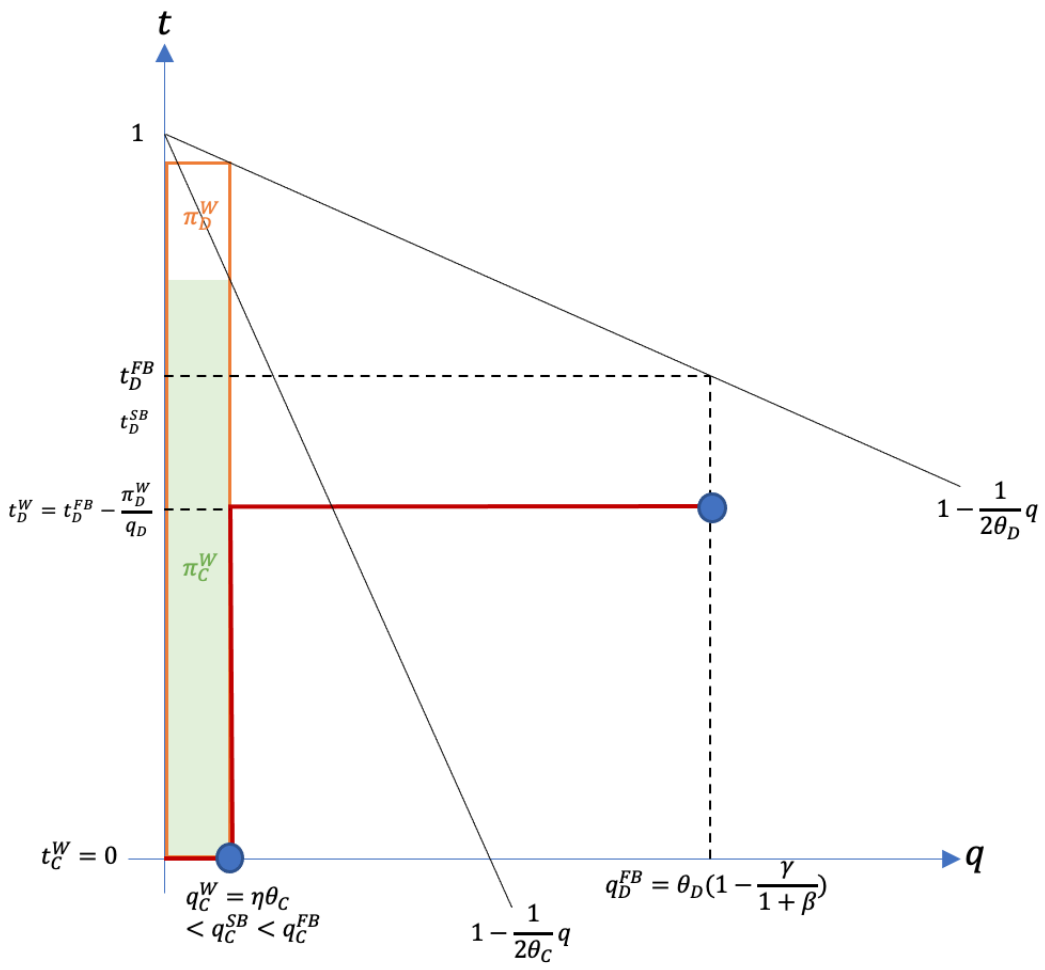


Figure 3: Regulation of Firms Located Abroad under Incomplete Information and WTO Rules

5 A Note on Firms' Choice of Technology

This section explores the interaction between inducing firms to choose the clean technology and its feasibility through an environmental policy under incomplete information.

We demonstrate that, as the "good" behavior (investing in the clean technology) leads to a less cost-efficient type (producing with the clean technology is more expensive than producing with the dirty technology), there is no way to incentivize firms to invest, as all firms prefer to stay dirty and pretend to have invested in the clean technology. It is straightforward to see this contradiction between truly

revealing technology and investing in the clean technology.

Using the same model as in previous sections, assume now that firms located abroad are initially all dirty firms with type θ_D. Before production, firms might choose to incur a R&D fixed cost $F > 0$ to acquire the clean technology of production (with type θ_C). For the sake of simplicity, the probability of success in developing the clean technology when performing the investment is equal to one.[15]

Assume it is valuable for society to induce firms to invest in clean technology, for example because environmental damage from emissions is infinitely large. Then, investing firms can be qualified as adopting the "good" behavior, as they choose to invest in a technology that creates no environmental damage.

If firms face the regulation described in Section 4 and are free to choose ex-ante their technology by deciding to incur the fixed cost of not, none of the firms would choose to invest in the clean technology, as clean firms already make lower profits compared to dirty firms ($\pi_C^W < \pi_D^W$).

The regulator wants to induce firms to invest in the clean technology, so he must adjust the policy to provide such incentive. Assume that we impose $t_C = 0$ as in Section 4. Therefore, the regulator would like to choose quantities q_D and q_C, a tax on dirty firms t_D, and a reward (subsidy) S to give to clean firms for investing F, in order to maximize society's welfare (1), such that firms perform the investment, reveal it to the regulator, and keep participating in the market. More specifically, the regulator must make sure that:

- It is more profitable for a firm to invest and become clean rather than staying dirty and revealing being dirty to the regulator:
  $q_C - \frac{1}{2\theta_C} q_C^2 - F + S \geq q_D - \frac{1}{2\theta_D} q_D^2 - t_D q_D$
- It is more profitable to invest and become clean rather than staying dirty and pretending to have invested: $q_C - \frac{1}{2\theta_C} q_C^2 - F + S \geq q_C - \frac{1}{2\theta_D} q_C^2 + S$

However, investing in the clean technology makes the firm less cost-efficient with $\theta_C < \theta_D$. Therefore, the latter condition can never be satisfied. To sum up, when dirty firms are more cost-efficient than clean firms, there is an incompatibility between inducing firms to choose the "good" behavior (investment in the clean technology) and inducing firms to reveal their true technology to the regulator (incentive compatibility constraints).

[15] Assuming that there exists a probability of failure in acquiring the clean technology only reinforces our point.

6 Conclusions

Since some regions regulate carbon emissions more stringently than others, global climate action is undermined by the issue of carbon leakage. A regulator may choose to regulate firms located abroad with a Carbon Border Adjustment Mechanism, aiming to impose a price on the carbon content of imported goods. However, it may be very costly, if not impossible, to acquire information about these firms' technologies of production (*i.e.* emissions). The absence of such information is the particular area of focus of this paper.

In a theoretical model, we investigate the design of a CBAM when firms located abroad keep as private information their technology of production. In order to properly distinguish between clean and dirty firms, the regulator must design a non-linear tax structure that differs from the domestic regulation, and thus induces a lower total welfare than under full information. However, this decrease in total welfare comes from a lower quantity produced by clean firms and a lower tax imposed on dirty firms. This implies that the mitigation of the environmental damage created by dirty firms is still optimal. The assumption that clean firms have a higher cost of production than dirty firms implies that, under incomplete information, the regulator is forced to choose a tax structure that rewards dirty firms and decreases the production of clean firms (although they create no environmental damage), for information purposes.

Including the WTO compliance constraints in a simple manner in the model sheds light on how international agreements may impose more distortions regarding border regulations compared to internal environmental policies.

This paper highlights an important friction between incentives to properly reveal technology and "good behavior" in the context of environmental issues. Indeed, when the "good" type (here, clean) is less cost-efficient than the "bad" type (dirty), none of the firms will voluntarily invest in the clean technology when facing this choice. If the regulator wants to construct a policy that incentivizes investment in the good type, it will inevitably clash with the incentive to properly reveal technology. This points to a key problem in the design of environmental policies under incomplete information, namely the potential impossibility of combining incentives to adopt a commendable behavior (e.g. investment in non-polluting production

processes) with incentives to report the true technology.

This paper is still a working paper, and additional research paths being explored are the following. We are considering adding consumers in the model, in order to include consumer surplus (and thus, price as well) in social welfare. This would allow for constructing a model where firms located abroad and regulated via the CBAM are foreign firms (and not domestic firms relocated abroad). This assumption changes the structure of the social planner's problem. Indeed, in this case, the regulator does not care about these firms' profits, but only about the tax revenue, environmental damage, and consumer surplus created by the consumption of the good. This setup can be considered without the inclusion of consumers, but when the regulator only cares about tax revenue and environmental damage, adding the WTO constraint implies zero production, making a version with consumer surplus more interesting.

References

Bohringer, C., E. J. Balistreri, and T. F. Rutherford. (2012). The role of border carbon adjustment in unilateral climate policy: overview of an Energy Modeling Forum study (EMF 29). Energy Economics.

Bohringer, C., Carbone, J. C., & Rutherford, T. F. (2018). Embodied carbon tariffs. The Scandinavian Journal of Economics, 120(1), 183-210.

Cosbey, A., Droege, S., Fischer, C., & Munnings, C. (2019). Developing guidance for implementing border carbon adjustments: Lessons, cautions, and research needs from the literature. Review of Environmental Economics and Policy, 13(1), 3-22.

Dasgupta, P., Hammond, P., & Maskin, E. (1980). On Imperfect Information and Optimal Pollution Control. The Review of Economic Studies, 47(5), 857-860.

Ellerman, A. D., Marcantonini, C., & Zaklan, A. (2016). The European Union emissions trading system: ten years and counting. Review of Environmental Economics and Policy, 10(1), 89-107.

European Union: European Commission. (2015). EU ETS Handbook.

European Commission, 2021. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Fit for 55': delivering the EU's 2030 Climate Target on the way to climate neutrality. COM/2021/550.

Hoel, M. (1996). Should a carbon tax be differentiated across sectors. Journal of Public Economics, 59, 17-32.

Kwerel, E. (1977). To Tell the Truth: Imperfect Information and Optimal Pollution Control. Review of Economic Studies. 44 (3), 595-601.

Laffont, J.-J., & Martimort, D. (2002). The Theory of Incentives: The Principal-Agent Model. Princeton University Press.

Laffont, J.-J., & Tirole, J. (1996). Pollution permits and environmental innovation. Journal of Public Economics, Elsevier, vol. 62(1-2), pages 127-140, October.

Markusen, J. (1975). International Externalities and Optimal Tax Structures. Journal of International Economics. 5. 15-29. 10.1016/0022-1996(75)90025-2.

Monjon, S., & Quirion, P. (2011). Addressing leakage in the EU ETS: Border adjustment or output-based allocation?. Ecological Economics, 70(11), 1957-1971.

Nordhaus, W. (2015). Climate clubs: overcoming free-riding in international climate policy. American Economic Review 105:1339-70.

Pauwelyn, J. (2013). Carbon leakage measures and border tax adjustments under WTO law. In Research handbook on environment, health and the WTO, ed. G. van Calster and D. Prevost, 448-506. Cheltenham: Edward Alger.

Pigou, A. (1920) The Economics of Welfare. MacMillan and Co., London.

Spulber, D. 1988. Optimal Environmental Regulation Under Asymmetric Information. Journal of Public Economics. 35(2): 163-181.

Stern, N. (2007). The Economics of Climate Change: The Stern Review. Cambridge: Cambridge University Press.

Stern, N., and J. E. Stiglitz. (2017). Report of the high- level commission on carbon prices. Washington, DC: World Bank Group.

Appendix

Proof of Proposition 1

The regulator solves the following problem:

$$\max_{q_D,q_C,T_C,T_D} W = \lambda\Big((1+\beta)\Big(q_C - \frac{1}{2\theta_C}q_C^2\Big) - \beta\pi_C\Big) + (1-\lambda)\Big((1+\beta)\Big(q_D - \frac{1}{2\theta_D}q_D^2\Big) - \gamma q_D - \beta\pi_D\Big) \quad (1.2)$$

subject to: $\pi_C(q_C) \geq 0$ and $\pi_D(q_D) \geq 0$.
Profits enter negatively in the welfare function, so the regulator chooses $\pi_C^{FB} = \pi_D^{FB} = 0$. To maintain these profits at zero, he has to offer, in the regulation, the following total taxes: $T_C^{FB} = \Big(q_C - \frac{1}{2\theta_C}q_C^2\Big)$ and $T_D^{FB} = \Big(q_D - \frac{1}{2\theta_D}q_D^2\Big)$. The solution of this problem with respect to quantities is given by deriving the first-order conditions:

- With respect to q_C: $q_C^{FB} = \theta_C$
- With respect to q_D: $q_D^{FB} = \theta_D\Big(1 - \frac{\gamma}{1+\beta}\Big)$

As a consequence; we can rewrite total taxes as their exact expressions:
$T_C^{FB} = \frac{1}{2}\theta_C$ and $T_D^{FB} = \frac{1}{2}\theta_D\Big(1 - \frac{\gamma^2}{(1+\beta)^2}\Big)$.

Proof of Proposition 2

The maximization problem of the regulator writes:

$$\max_{q_D,q_C,T_C,T_D} W = \lambda\Big((1+\beta)\Big(q_C - \frac{1}{2\theta_C}q_C^2\Big) - \beta\pi_C\Big) + (1-\lambda)\Big((1+\beta)\Big(q_D - \frac{1}{2\theta_D}q_D^2\Big) - \gamma q_D - \beta\pi_D\Big) \quad (1.2)$$

subject to:

$$q_C - \frac{1}{2\theta_C}q_C^2 - t_C q_C \geq q_D - \frac{1}{2\theta_C}q_D^2 - t_D q_D \quad (2)$$

$$q_D - \frac{1}{2\theta_D}q_D^2 - t_D q_D \geq q_C - \frac{1}{2\theta_D}q_C^2 - t_C q_C \quad (3)$$

$$\pi_C(q_C, t_C) \geq 0 \quad (4)$$

$$\pi_D(q_D, t_D) \geq 0 \quad (5)$$

First, note that we need $q_C < q_D$ for both (2) and (3) to hold simultaneously. As discussed in the previous subsection, the problem of incentives usually arises from dirty technology firms, so we can ignore the constraint ensuring that the clean type behave according to their true technology (2) and check that it is satisfied ex-post. Constraint (3) can be rewritten $\pi_D \geq \pi_C + q_C^2\Big(\frac{1}{2\theta_C} - \frac{1}{2\theta_D}\Big)$. Constraint (4) imposes non-negative profits for clean firms, and

the second term on the right-hand side is strictly positive, so the participation constraint of the dirty type (5) is necessarily satisfied and can be ignored. Because π_C and π_D enter negatively in the objective welfare function (1), the regulator would like to set them at their lowest value possible. As a consequence, both remaining constraints hold with equality: the regulator chooses $\pi_C = 0$ and $\pi_D = q_C^2\left(\frac{1}{2\theta_C} - \frac{1}{2\theta_D}\right)$. Proceeding by substitution, the simplified maximization problem rewrites:

$$\max_{q_D, q_C \geq 0} \lambda\left((1+\beta)\left(q_C - \frac{1}{2\theta_C}q_C^2\right)\right) + (1-\lambda)\left((1+\beta)\left(q_D - \frac{1}{2\theta_D}q_D^2\right) - \gamma q_D - \beta q_C^2\left(\frac{1}{2\theta_C} - \frac{1}{2\theta_D}\right)\right)$$

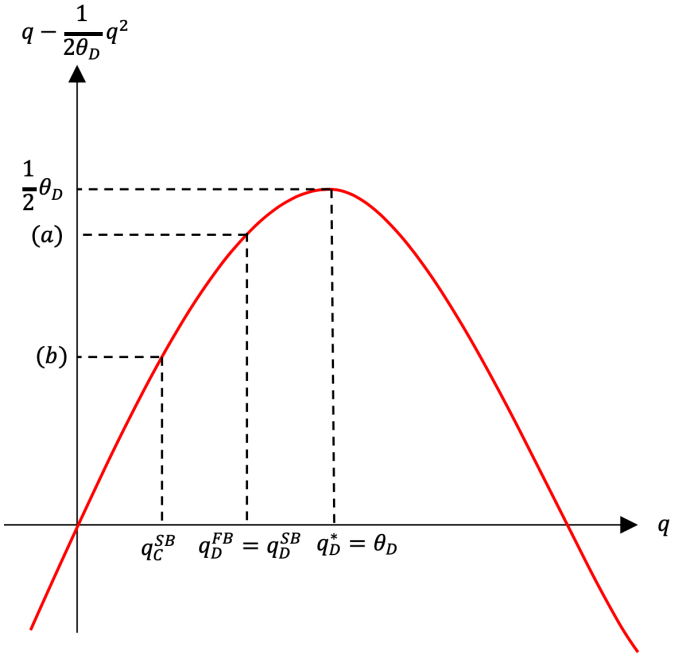
Computing the first-order conditions and using the constraints give the following solutions:

- $q_D = \theta_D\left(1 - \frac{\gamma}{1+\beta}\right)$
- $q_C^{SB} = \theta_C \frac{\theta_D \lambda (1+\beta)}{\theta_D(\lambda+\beta) - \theta_C \beta (1-\lambda)}$
- $\pi_D^{SB} = q_C^2\left(\frac{1}{2\theta_C} - \frac{1}{2\theta_D}\right)$
- $\pi_C^{SB} = 0$
- $T_D^{SB} = T_D^{FB} - \pi_D^{SB}$, so $t_D^{SB} = t_D^{FB} - \frac{\pi_D^{SB}}{q_D}$
- $T_C^{SB} = q_C^{SB} - \frac{1}{2\theta_C} q_C^{SB^2}$, so $t_C^{SB} = 1 - \frac{1}{2\theta_C} q_C^{SB}$

Showing that $t_D^{SB} < t_D^{FB}$ is straightforward as $\pi_D^{SB} > 0$ and $q_D > 0$. Showing that $T_C^{SB} < T_D^{SB}$ is straightforward as well with a graphical illustration. Indeed, replacing the expressions in T_D^{SB}, we can rewrite:

$$T_D^{SB} = T_C^{SB} + \underbrace{q_D^{SB}\left(1 - \frac{1}{2\theta_D} q_D^{SB}\right)}_{(a)} - \underbrace{q_C^{SB}\left(1 - \frac{1}{2\theta_D} q_C^{SB}\right)}_{(b)}.$$

The following graphical illustration shows that $(a) > (b)$.



Proof of Proposition 3

Sketch of proof - Footnote 12. This reasoning aims to provide intuition for the validity of the following statement: *"For all values of* $\beta \in [0,1]$*, there exists a range of acceptable values of* γ *that satisfy the WTO requirement under second-best taxation, i.e.* $t_D^{SB} \leq \gamma$*"*.

Recall, in first-best we have $t_D^{FB} = \frac{1}{2}\left(1 + \frac{\gamma}{1+\beta}\right)$. The first-best per-unit tax would satisfy $t_D^{FB} \leq \gamma$ if and only if $\gamma \geq \frac{1+\beta}{1+2\beta}$, where we also assume $\gamma < 1 + \beta$ for non-negativity of q_D. Therefore, $t_D^{FB} \leq \gamma$ is satisfied for all values of $\gamma \in \left[\frac{1+\beta}{1+2\beta}; (1+\beta)\right]$. Therefore, there exists values of γ inducing the first-best per-unit tax on dirty firms to satisfy the WTO rule. Now, we know that $t_D^{SB} < t_D^{FB}$. As a direct consequence, we can state that there also exists acceptable values of γ induxing the second-best per-unit tax on dirty firms to satisfy the WTO rule, *i.e.* $t_D^{SB} \leq \gamma$.

Simplifying and solving the problem. The maximization problem of the regulator writes:

$$\max_{q_D, q_C, T_C, T_D} W = \lambda\left((1+\beta)\left(q_C - \frac{1}{2\theta_C}q_C^2\right) - \beta\pi_C\right) + (1-\lambda)\left((1+\beta)\left(q_D - \frac{1}{2\theta_D}q_D^2\right) - \gamma q_D - \beta\pi_D\right) \quad (1.2)$$

subject to:

$$\pi_C \geq q_D - \frac{1}{2\theta_C} q_D^2 - t_D q_D \quad (6)$$

$$\pi_D \geq q_C - \frac{1}{2\theta_D} q_C^2 \quad (7)$$

$$\pi_C \geq 0 \quad (8)$$

$$\pi_D \geq 0 \quad (9)$$

$$t_C = 0 \Rightarrow \pi_C = q_C - \frac{1}{2\theta_C} q_C^2 \quad (10)$$

$$t_D \leq \gamma \quad (11)$$

We ignore (6) and check that it is satisfied ex-post, as the problem of incentives arises from dirty firms and not clean firms. We also ignore (11) and check that there exists values of the marginal damage γ such that it holds ex-post. Given (10), then constraint (8) is satisfied for any $q_C \in [0, 2\theta_C]$, which we check ex-post as well. Of constraints (7) and (9), only one can be binding. Therefore, one of them is irrelevant in the program. We can show that (9) is irrelevant and (7) is binding at the optimum. Proceed by contradiction. If (9) binds before (7), then $\pi_D = 0$. In this case, solving the program yields $q_C = \theta_C$. But this would lead to dirty firms making positive profits when reporting θ_C (lie), with $\pi_D(q_C, T_C) = \theta_C\big(1 - \frac{\theta_C}{2\theta_D}\big) > 0$. Therefore, the constraint on dirty firms to make sure they tell the truth about their technology (7) is binding at the optimum, and we have $\pi_D = q_C - \frac{1}{2\theta_D} q_C^2$. Proceeding by substitution, the simplified maximization problem rewrites:

$$\max_{q_D, q_C} W = \lambda\Big(q_C - \frac{1}{2\theta_C} q_C^2\Big) + (1-\lambda)\Big[(1+\beta)\Big(q_D - \frac{1}{2\theta_D} q_D^2\Big) - \gamma q_D - \beta\Big(q_C - \frac{1}{2\theta_D} q_C^2\Big)\Big]$$

Computing the first-order conditions and using the constraints give the following solutions:

- $q_D = \theta_D\big(1 - \frac{\gamma}{1+\beta}\big)$
- $q_C^W = \theta_C \frac{\theta_D(\lambda - (1-\lambda)\beta)}{\lambda\theta_D - (1-\lambda)\beta\theta_C}$
- $\pi_D^W = q_C^W\big(1 - \frac{1}{2\theta_D} q_C^W\big)$
- $\pi_C^W = q_C^W\big(1 - \frac{1}{2\theta_C} q_C^W\big)$
- $T_D^W = T_D^{FB} - \pi_D^W$, so $t_D^W = t_D^{FB} - \frac{\pi_D^W}{q_D}$
- $t_C^W = 0$

With these results, we can check ex-post the ignored constraints. Participation constraint for clean firms (3) is satisfied for q_C^W. Intuitively, there exists an interval of acceptable values of γ for which $t_D^W \leq \gamma$, as we have $t_D^W < t_D^{FB}$ (refer to the sketch of proof of footnote 12 for a similar reasoning). Therefore, constraint (11) can be satisfied ex-post.

We want to prove that $q_C^W < q_C^{SB}$. By replacing each quantity by its expression, we get $\beta^2(1-\lambda)(\theta_D - \theta_C) > 0$. This is true as long as $\theta_D > \theta_C$, which holds as an assumption of the model.

Now compare the profits given up to dirty firms in second-best with respect to WTO. Recall, we can write: $q_C^{SB} = \varphi q_C^{FB}$ where $\varphi = \dfrac{\theta_D \lambda(1+\beta)}{\theta_D(\lambda+\beta) - \theta_C\beta(1-\lambda)} < 1$ and $q_C^W = \eta q_C^{FB}$ where $\eta = \dfrac{\theta_D(\lambda - (1-\lambda)\beta)}{\lambda\theta_D - (1-\lambda)\beta\theta_D} < \varphi < 1$. So replacing gives: $\pi_D^{SB} = \theta_C^2\varphi^2\Big(\dfrac{1}{2\theta_C} - \dfrac{1}{2\theta_D}\Big)$ and $\pi_D^W = \theta_C\eta\Big(1 - \dfrac{1}{2\theta_D}\theta_C\eta\Big)$.

Study the conditions under which we have $\pi_D^W > \pi_D^{SB}$:

$$\theta_C\eta\Big(1 - \frac{1}{2\theta_D}\theta_C\eta\Big) > \theta_C^2\varphi^2\Big(\frac{1}{2\theta_C} - \frac{1}{2\theta_D}\Big) \;\Leftrightarrow\; \eta - \frac{\theta_C}{2\theta_D}\eta^2 - \frac{1}{2}\varphi^2 + \frac{\theta_C}{2\theta_D}\varphi^2 > 0$$

First assume that η and φ are close. For simplification, look at the case where $\eta = \varphi$. This happens when β tends towards zero and/or λ tends towards 1. The inequality becomes $\dfrac{-1}{2}\varphi^2 + \varphi > 0$ which holds for any $\varphi \in [0, 2]$, therefore is always satisfied. Thus, for η and φ close (*i.e.* β small and/or λ high), we have $\pi_D^W > \pi_D^{SB}$. Now consider an increase in β and/or a decrease in λ. Given that $0 < \eta < \varphi < 1$, we can do the latter until $\eta = 0 \Leftrightarrow \lambda - (1-\lambda)\beta = 0$. In this case, the LHS of the inequality becomes: $\dfrac{-1}{2}\varphi^2 + \dfrac{\theta_C}{2\theta_D}\varphi^2$. With the assumption that $\theta_D > \theta_C$, this expression is negative. Therefore, when β is high enough and/or λ small enough, the inequality is reversed and $\pi_D^W < \pi_D^{SB}$. This analysis translates directly into comparing the total and per-unit taxes imposed on dirty firms in second-best compared to WTO.